

09/163848

ABSTRACT

A method and apparatus for automatic document classification using text and images. The present invention provides a method and apparatus for automatic document classification based on text and image. A new document is analyzed based on textual content as well as visual appearance. The new document is automatically stored in one or more mirror directories in which the new document would most likely be stored by the user of the device if the new document were placed manually. Determination of the most likely directories is based on an analysis of multiple documents stored by the user in various directories. The mirror directories are components of a mirror directory structure, which is a copy of a pre-existing directory structure, such as the user's hard drive. By storing the new document automatically, the user is relieved of the duty of manually selecting a directory for the new document.